Genomics 98 (2011) 376–380

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

phiGENOME: An integrative navigation throughout bacteriophage genomes

Matej Stano, Lubos Klucar *

Laboratory of Bioinformatics, Institute of Molecular Biology, Slovak Academy of Sciences, Dubravska cesta 21, Bratislava, Slovakia

ARTICLE INFO

Article history:
Received 4 April 2011
Accepted 21 July 2011
Available online 30 July 2011

Keywords:
Bacteriophage
Genome browser
Genome visualisation
Gene regulation

ABSTRACT

phiGENOME is a web-based genome browser generating dynamic and interactive graphical representation of phage genomes stored in the phiSITE, database of gene regulation in bacteriophages. phiGENOME is an integral part of the phiSITE web portal (http://www.phisite.org/phigenome) and it was optimised for visualisation of phage genomes with the emphasis on the gene regulatory elements. phiGENOME consists of three components: (i) *genome map viewer* built using Adobe Flash technology, providing dynamic and interactive graphical display of phage genomes; (ii) *sequence browser* based on precisely formatted HTML tags, providing detailed exploration of genome features on the sequence level and (iii) *regulation illustrator*, based on Scalable Vector Graphics (SVG) and designed for graphical representation of gene regulations. Bringing 542 complete genome sequences accompanied with their rich annotations and references, makes phiGENOME a unique information resource in the field of phage genomics.

1. Introduction

Bacteriophages (aka phages) are the most abundant biological entities on Earth [1]. Regarding their world-wide population (estimated to be 10^{31}) and remarkable genetic diversity, phages likely represent the most plentiful repository of unexplored genes [2]. Phages were the primary biological entities studied on a molecular level and the first with complete genomes to be determined [3,4]. Nowadays, above 600 complete phage genome sequences are known (phage section of the NCBI Genome database contains 644 completely sequenced genomes, http://www.ncbi.nlm.nih.gov/genome). Despite this number, there is a lack of biological databases addressing specific features of phages and their genomes. One of the few is phiSITE — database of gene regulation in bacteriophages [5].

The phiSITE was primarily developed in order to collect and describe transcription regulators in phage genomes. To ensure that users will retrieve information efficiently, we decided to integrate suitable graphical genome browsing module to the phiSITE web interface. Initial survey for existing genome browsers revealed several software packages available for local installation and implementation for custom projects. Three of them, Ensembl [6], UCSC Genome Browser [7] and GBrowse [8], are extensively used by the scientific community, e.g. within the Model organism databases (MODs) [9]. Ensembl [6] (http://www.ensembl.org) provides online access to the chordate genome sequences and annotation with a particular focus on human, mouse, rat, zebrafish and other high-value sequenced genomes. Additional specific resources are addressed to the genome variation, function and evolution. The University of California, Santa Cruz (UCSC) Genome Browser [7] (http://genome.ucsc.edu) integrates genomic information for the human genome and dozens of other vertebrate, invertebrate and yeast genomes. Incorporated tools enable examination and comparative analyses of genomes, building of sequence alignments and display of users' own annotation data. The Generic Genome Browser [8] (GBrowse; http://www.gbrowse.org) is a web-based application for displaying genomic annotation and other features. Browser functionalities include the ability to scroll and zoom through arbitrary regions of a genome, searching for a landmark or performing a full text search and providing customizable output. Regarding the field of phage genomics, GBrowse is included e.g. in T4-like genome database [10]. Worth mentioning is also NCBI Map Viewer [11], which facilitates browsing of all biological sequences stored in the NCBI databases. However in contrary with the previously listed tools, it is not customizable and cannot be installed locally. Although originally developed with different focus, basically all referred genome browsers are optimised for visualisation and analysis of large eukaryotic genomes. Current abundance of genome browsers results from variety of needs they are expected to meet and hence it is unlikely to develop one universal genome browser [8].

Subsequently, we explored genome display capabilities in databases of gene regulation in prokaryotes which store data similar to those in the phiSITE database. Databases PRODORIC [12], RegulonDB [13] and CoryneRegNet [14] — they all provide graphical representations in the form of image maps. Although dynamically generated, they represent a server-side technology based on hyperlinked bitmaps, hence the dynamics and interactivity are limited.

We were demanding genome visualisation in a dynamic, interactive, lucid and graphically subtle way with emphasised display of *cis*-regulatory elements and their regulatory interactions. Since none of the

* Corresponding author. Fax: +421 2 5930 7416.
E-mail addresses: matej.stano@savba.sk (M. Stano), lubos.klucar@savba.sk (L. Klucar).

0888-7543/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2011.07.004

available applications and technologies was able to fulfil all these requirements for phiISTE genome browsing module, we initiated development of a novel, custom-made application phiGENOME, described in this paper.

2. Material and methods

2.1. Data content

Genome annotation data stored in the phiSITE database are extracted from EMBL entries and loaded into database using semi-automatic PHP-based parsing utility. Specific information on *cis*-regulatory sites and regulatory interactions are acquired primarily from scientific papers and books and are deposited into the database manually. Detailed description of the phiSITE architecture and content is described by Klucar et al. [5]. To date, the database contains 542 annotated complete phage genomes from 16 phage families and covers more than 42,000 genes (Table 1). Besides genes, annotations for *cis*-regulatory elements (promoters, operators, terminators and attachment sites) from 33 phages are included. Comprehensive gene regulatory networks of mutual regulatory interactions between *cis*-regulatory elements and genes are currently defined for 9 phages where gene regulation was well-studied. These networks are constructed using the BioTapestry, a Java based application for building, visualisation and simulation of genetic regulatory networks [15].

2.2. Development

phiGENOME is integrated into the phiSITE database framework, which was built on well-proven LAMP platform (Linux, Apache, MySQL, PHP). Graphical output of phiGENOME is generated dynamically; required data are retrieved on request from the MySQL back-end and the data transfer is facilitated via XML files.

The key feature of phiGENOME is the Flash-based *genome map viewer*. We adopted Adobe Flash technology (http://www.adobe.com/flashplatform) considering its rendering abilities, efficient vector graphics, versatility and cross-platform independence. The phiGENOME Flash applet was developed in ActionScript 3.0, using freely available SEPY ActionScript editor (http://www.sephiroth.it/python/sepy.php). SWF binaries were compiled with a free Flex compiler from Adobe Flex 3.0 SDK (http://flex.org).

Sequence browser and regulation illustrator, the other components of the phiGENOME, are server-side applications built on PHP 5.2.2. (http://www.php.net). Semi-graphical output of the *sequence browser* is based on precisely formatted HTML tags, whereas output of *regulation illustrator* is based on Scalable Vector Graphics (SVG) markup language version 1.1 (http://www.w3.org/Graphics/SVG/).

The cross-communication between all three components of the phiGENOME is facilitated via the set of JavaScript functions.

Table 1
phiGENOME browser coverage (phage families breakdown).

Nucleic acid	Phage family	Number of phages
dsDNA	Siphoviridae	239
	Myoviridae	101
	Podoviridae	89
	Lipothrixviridae	8
	Fuselloviridae	8
	Tectiviridae	4
	Rudiviridae	4
	Bicaudaviridae	2
	Globuloviridae	2
	Plasmaviridae	1
	Corticoviridae	1
	Ampullaviridae	1
ssDNA	Inoviridae	25
	Microviridae	14
dsRNA	Cystoviridae	5
ssRNA	Leviviridae	7
N/A	unclassified phages	31
		542

3. Results

3.1. Overview

phiGENOME is a web-based genome browsing module for interactive navigation throughout the phage genomes stored in the phiSITE database. The size of phage genomes spans a wide range from the smallest Leuconostoc phage L5 genome (2435 bp) to the largest Pseudomonas phage 201phi2-1 with its genome size of 316,674 bp. Regarding the composition (DNA/RNA; single-stranded/double-stranded) and topology (linear/circular/segmented), phage genomes (like all viral genomes) are largely diverse. Since phiGENOME was optimised for phage genome display, it is able to cope with all these peculiarities.

Search through phage collection is facilitated by Ajax auto-complete search script. Phages with annotated *cis*-regulatory elements are highlighted in the selection menu with a light blue background. When the user selects phage of his/her interest, phiGENOME user's interface appears. This consists of the three components: (i) *genome map viewer* built on the Adobe Flash technology (Fig. 1A), (ii) *sequence browser* based on precisely formatted HTML tags (Fig. 1B) and (iii) *regulation illustrator* employing SVG (Fig. 2). All components are closely coupled, thus any action performed in one of them is immediately projected in the other one. Accordingly, the sequence displayed in the *sequence browser* matches the genome segment displayed in the *genome map viewer*. Selecting an element in any graphical component makes it selected also in the two remaining components. By default, all components of the phiGENOME user's interface are enabled. Optional disabling of *sequence browser* and *regulation illustrator* makes the browsing less data-intensive.

3.2. Genome map viewer

The *genome map viewer* (Fig. 1A) provides highly dynamic graphical display of the phage genome. Both linear and circular genomes are displayed in the linear mode. The strand of genome molecule is represented as a plain grey bar. If the genome is double-stranded both strands are displayed alongside each other. In the case of single-stranded genomes, where genes are coded in both directions, the complementary strand is depicted in lighter shade (e.g. Vibrio phage VSK genome). Particular genomic features (genes, promoters, operators, terminators, attachment sites) are depicted as semi-transparent coloured boxes matching their real location and size in a genome. Each type of genetic feature is assigned with a specific colour (genes are blue; promoters are red; operators are green; terminators are yellow and attachment sites are purple), following the colour schema common for the whole phiSITE web-interface. Semitransparency of boxes facilitates display of overlapping genes and regulatory elements, which are quite frequent in small phage genomes. In the upper left side of the *genome map viewer* navigation bar is located (or navigation circle in the case of circular genome), which schematically represents the entire genome and indicates the position of currently displayed genome segment. In the upper right corner, zoom indicator is located, which shows the level of the magnification for displayed genome map. The initial display in *genome map viewer* spans an entire genome. Mouse scroll can be used to change the zoom for genome maps in order to perform detailed inspection and to recognise closely spaced features. Whilst zooming, neighbouring feature labels progressively appear and disappear in

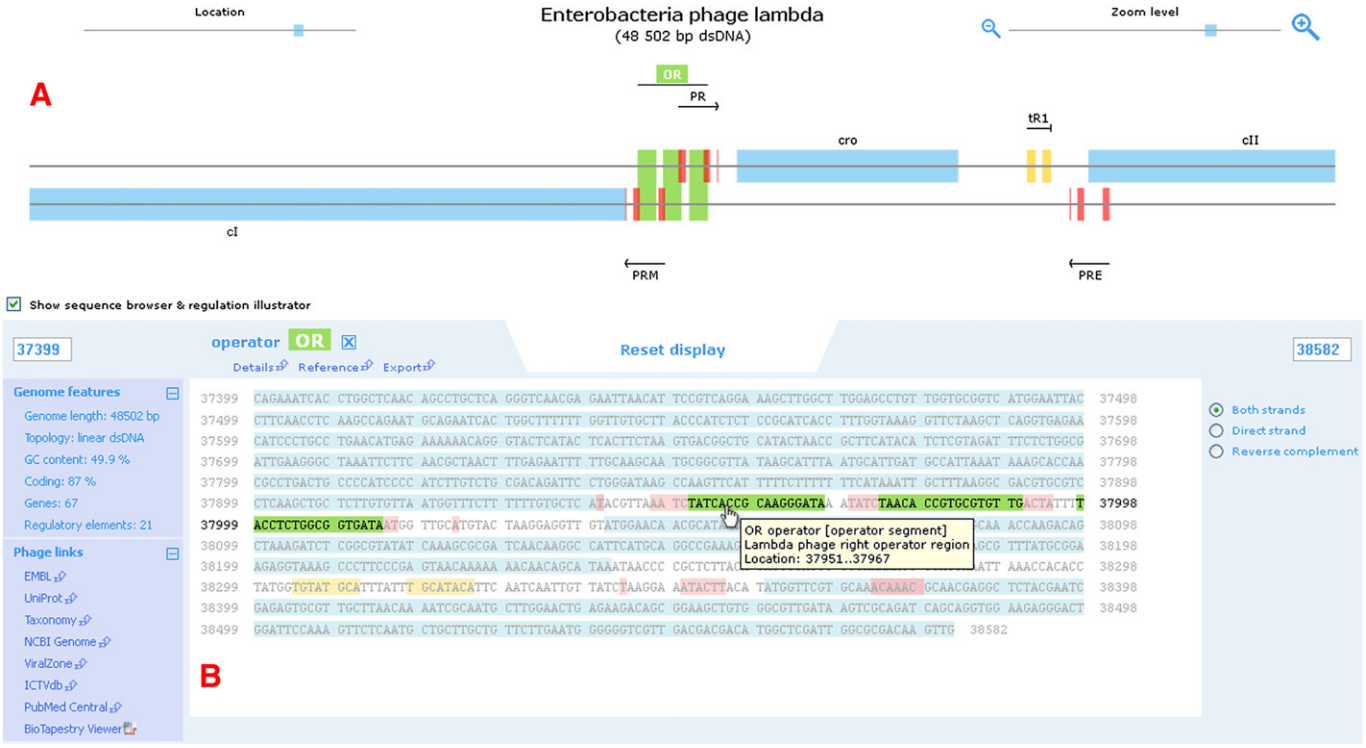

Fig. 1. phiGENOME browser displaying Enterobacteria phage lambda lysis/lysogeny switching region. (A) Flash-based *genome map viewer*. Selected genomic feature (operator OR) has highlighted label. On the left-hand top corner is located navigation bar indicating position of currently displayed genome segment. In the upper right side, the zoom indicator with zoom controllers is situated. (B) HTML-based *sequence browser*. Selected operator (OR) is highlighted; the rest of the sequence is displayed in semi-transparent mode. On the left, general features of the phage genome are listed and accompanied with menu of associated hypertext links. Radio buttons on the right-hand side are used to define strand whose features should be displayed.

order to keep them readable. At a specific magnification level, actual nucleotide sequence shows up and replaces the grey bar representation of the strand. Supplementary zooming functionality is enabled with "+" and "−" buttons and by sliding the zoom level indicator. The easiest way to move along genome map is to use a drag-and-drop technique, where user can move a visible part of the genome by dragging any genome element. In addition, user can view desired genome region by submitting required begin and end coordinates manually into the text boxes. Mouse over a feature displays pop-up tooltip containing basic information about the feature (feature's description, type and location). Selecting arbitrary feature by mouse click makes this feature highlighted and enables set of cross-referencing links associated with the feature. These interconnect phiGENOME with other sections of the phiSITE database web interface (search for regulatory sites, data export in FASTA or XML format), with external databases (EMBL, UniProt, PubMed) and with NCBI BLAST Search. All these functions make browsing more intuitive and interactive.

3.3. Sequence browser

Sequence browser (Fig. 1B), the other component of the phiGENOME, affords detailed exploration of genome features on the sequence level. Genome features are highlighted in colours which follow the colour scheme used in the Flash graphics (including blended colours of overlapping regions). Mouse over and mouse click events have the same effect as in the *genome map viewer*. Unlike *genome map viewer*, only single strand is displayed in the *sequence browser*, either direct (default setting) or reverse-complement. User may choose whether genome features from a single strand or from both superposed strands will be displayed. To the left from the division of *sequence browser*, two expandable menus are placed. One of them covers general properties of the displayed phage genome (length, topology and composition, percentage of GC content, percentage of coding sequences, number of genes and number of *cis*-regulatory elements, if defined). The second menu contains links to external databases (EMBL, UniProt, NCBI Genome, NCBI Taxonomy, ViralZone, ICTVdB, PubMed Central) associated with the displayed phage genome. If the gene regulatory network is defined for the phage, link to the BioTapestry model is also available.

3.4. Regulation illustrator

The third part of the phiGENOME was designed for graphical representation of regulatory interactions between genes and *cis*-regulatory elements. For any component of defined regulatory network, *regulation illustrator* (Fig. 2) depicts the closest regulatory neighbours in the form of directed acyclic graph. Vertices (icons) stand for individual components of the regulatory network and edges (arrows) represent interactions between them. Selected element is represented by the central icon. Elements influencing this central element are located on the left, elements influenced by this element are located on the right. Each type of genetic feature is represented by specific graphical icon coloured according to the phiSITE standard colour scheme. The nature of regulations can be positive (represented by green-coloured regular arrows), negative (red T-shaped arrows) or neutral (grey lines without arrows). Mouse over an icon displays tooltip with basic description of the element. Mouse over an arrow shows tooltip text containing the type of interaction. Mouse click on any icon generates graph of regulations for this element. User has an option to download PNG snapshot of any regulatory scheme. Besides *regulation illustrator*, user may explore comprehensive regulatory networks using BioTapestry Viewer, where particular subsets of regulatory network are divided into separate graphical layers. These regulatory subsets are based on the function of the genes, their regulation and time of expression.

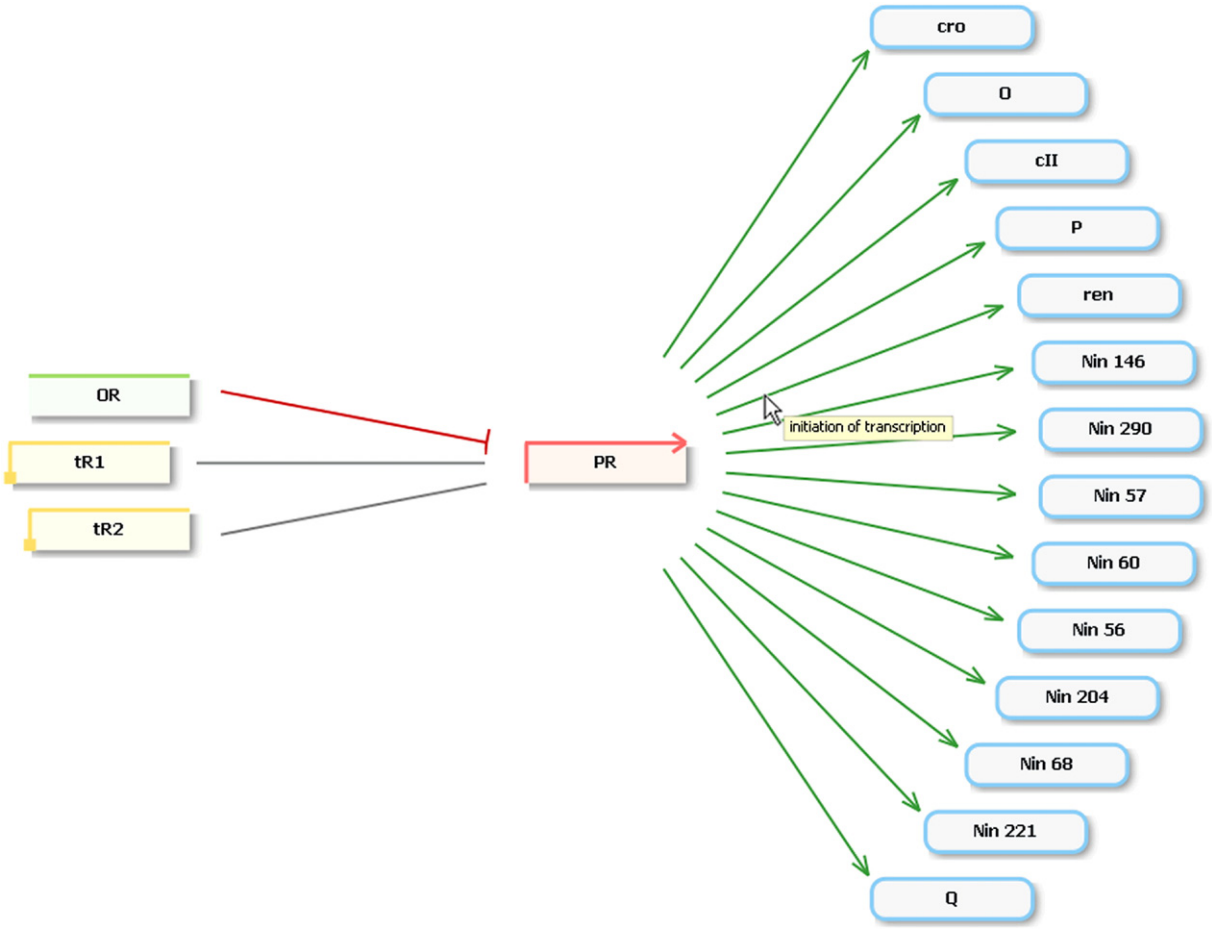

Fig. 2. Visualisation of regulatory interactions for Enterobacteria phage lambda promoter PR. The PR promoter icon represents the central element of the graphical scheme. Operator OR (green icon) represses transcription from this promoter. Terminators tR1 and tR2 (yellow icons) terminate transcription from promoter PR. There are 14 genes on the right (blue icons) whose transcription is initiated from promoter PR.

3.5. Availability and requirements

phiGENOME is accessible at the phiSITE database web-portal: http://www.phisite.org/phigenome. It is freely accessible to any individual and for any purpose and it is distributed under the "Creative Commons Attribution-Share Alike 3.0 Unported License" (http://creativecommons.org/licenses/by-sa/3.0/). An Adobe Flash Player plug-in 9.0 or higher is required to be installed into user's web browser.

For web browsers lacking the native SVG support (e.g. Microsoft Internet Explorer ver. 8 or lower), installation of SVG plug-in is necessary.

4. Discussion

phiGENOME is a web-based genome browser generating dynamic and interactive graphical representation of bacteriophage genomes stored in the phiSITE database. The first publically accessible version of the phiSITE genome browser was described by Klucar et al. (2010) [5]. This preliminary version was incomplete and immature and lacked most of the features present in the phiGENOME browser described herein. phiGENOME (current version 1.4) is the official and well-proven release of the genome browser which encompasses many substantial improvements. The content of the phiSITE database was significantly upgraded. The number of phage genome entries has increased from few tens to the current number of 542 annotated genomes. Eventually, our goal is to expand phiSITE to a comprehensive repository of phage regulatory sequences, genes and genomes and to make phiGENOME a universal gateway to the phage knowledge universe. For the further development it is worth considering including also prophage entries, though their annotation is usually very limited.

Numerous changes affected also the user's interface. Novel modules, *sequence browser* and *regulation illustrator*, were developed to complement Flash graphics. *Sequence browser* facilitates browsing through genome features on a sequence level; *regulation illustrator* complements genome browser with information focused on gene regulation.

The Flash applet itself was re-designed and improved in many ways. Preliminary version of genome browser lacked interactivity and cross-referencing with other modules of the phiSITE web portal. Addition of numerous references to external resources significantly enhanced its interactivity. Advanced graphical and navigation functions like mouse-wheel zooming, the navigation bar and dynamic feature labels promote intelligibility and user-friendliness of the phiGENOME. Rendering of genomic features and nucleotide sequence is more precise. Nevertheless, the limited length of phage genomes has facilitated implementation of several graphical features that would be more difficult (or impossible) to achieve with longer and more complex genomes or chromosomes. Additionally, display abilities were upgraded and adjusted for visualisation of any phage genome, regardless of the composition and topology.

Adobe Flash was chosen as a major technology for development of phiGENOME graphics. At the time, other technologies that could substitute Flash in the future (e.g. HTML5) are not yet mature and widely used. Flash programming platform is optimised for development of dynamic and responsive web applications and the usage of

graphical elements is highly efficient. The size of Flash applets is quite small (125 kB for phiGENOME), thus they are fast to download (e.g. comparing to JAVA applets). Although Flash Player plug-in is necessary for phiGENOME to be functional in a web browsers, dominant penetration of Flash technology throughout the Internet and ubiquitous support by all modern browsers make it available on most user-end systems (phiSITE Google Analytics survey shows 96% penetration of Flash-enabled web browsers). A major drawback is unequal performance of Flash applications on various operation systems. As a result, e.g. phiGENOME running on Mac OS platforms with older version of Flash Player (<10.1) has unavailable mouse scroll zoom. Although Adobe Flash is a proprietary software, ActionScript source files can be compiled using freely available Adobe Flex compiler. This makes Flash technology more accessible to all potential developers and allows its wider usage also in academic communities.

phiGENOME is an innovative application, considering the mode and purpose how the Flash technology was employed in phage genomic context. Nevertheless, its main purpose is strictly biological and it is intended to be used by a wide range of researchers studying bacteriophages on different levels. It is a unique resource providing insights into the phage genomics and gene regulation with a solid depth of knowledge accompanied with a user friendly interface.

5. Conclusions

One of the essential challenges in the field of Bioinformatics is to facilitate access to biological data and knowledge. phiGENOME represents unique, publically available resource integrating all available information related to bacteriophage genomics and gene regulation. It provides intelligible and visually sophisticated access to the information on bacteriophage *cis*-regulatory elements, genes and genomes collected in the phiSITE database. The value of retrieved information is enhanced by interactive references to related internal and external resources.

Acknowledgments

We thank Matus Hajduk and Diana Hierwegova for literature mining related to regulatory interactions in phages and subsequent building of the gene regulatory networks. We are also grateful to Renata Novakova for her helpful comments and suggestions and Katarina Hojckova for the careful reading of the manuscript. This work was supported by the Slovak Research and Development Agency [grant number APVT-51-025004]; and the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and of the Slovak Academy of Sciences [grant number VEGA 2/0100/09].

References

[1] R.W. Hendrix, Bacteriophage genomics, Curr. Opin. Microbiol. 6 (2003) 506–511.
[2] G.F. Hatfull, Bacteriophage genomics, Curr. Opin. Microbiol. 11 (2008) 447–453.
[3] F. Sanger, G.M. Air, B.G. Barrell, N.L. Brown, A.R. Coulson, C.A. Fiddes, C.A. Hutchison, P.M. Slocombe, M. Smith, Nucleotide sequence of bacteriophage phi X174 DNA, Nature 265 (1977) 687–695.
[4] W. Fiers, R. Contreras, F. Duerinck, G. Haegeman, D. Iserentant, J. Merregaert, W. Min Jou, F. Molemans, A. Raeymaekers, A. Van den Berghe, G. Volckaert, M. Ysebaert, Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene, Nature 260 (1976) 500–507.
[5] L. Klucar, M. Stano, M. Hajduk, phiSITE: Database of gene regulation in bacteriophages, Nucleic Acids Res. 38 (2010) D366–D370.
[6] P. Flicek, M.R. Amode, D. Barrell, K. Beal, S. Brent, Y. Chen, P. Clapham, G. Coates, S. Fairley, S. Fitzgerald, L. Gordon, M. Hendrix, T. Hourlier, N. Johnson, A. Kähäri, D. Keefe, S. Keenan, R. Kinsella, F. Kokocinski, E. Kulesha, P. Larsson, I. Longden, W. McLaren, B. Overduin, B. Pritchard, H.S. Riat, D. Rios, G.R. Ritchie, M. Ruffier, M. Schuster, D. Sobral, G. Spudich, Y.A. Tang, S. Trevanion, J. Vandrovcova, A.J. Vilella, S. White, S.P. Wilder, A. Zadissa, J. Zamora, B.L. Aken, E. Birney, F. Cunningham, I. Dunham, R. Durbin, X.M. Fernández-Suarez, J. Herrero, T.J. Hubbard, A. Parker, G. Proctor, J. Vogel, S.M. Searle, Ensembl 2011, Nucleic Acids Res. 39 (2011) D800–D806.
[7] P.A. Fujita, B. Rhead, A.S. Zweig, A.S. Hinrichs, D. Karolchik, M.S. Cline, M. Goldman, G.P. Barber, H. Clawson, A. Coelho, M. Diekhans, T.R. Dreszer, B.M. Giardine, R.A. Harte, J. Hillman-Jackson, F. Hsu, V. Kirkup, R.M. Kuhn, K. Learned, C.H. Li, L.R. Meyer, A. Pohl, B.J. Raney, K.R. Rosenbloom, K.E. Smith, D. Haussler, W.J. Kent, The UCSC Genome Browser database: update 2011, Nucleic Acids Res. 39 (2011) D876–D882.
[8] L.D. Stein, C. Mungall, S. Shu, M. Caudy, M. Mangone, A. Day, E. Nickerson, J.E. Stajich, T.W. Harris, A. Arva, S. Lewis, The generic genome browser: a building block for a model organism system database, Genome Res. 12 (2002) 1599–1610.
[9] T.Z. Sen, L.C. Harper, M.L. Schaeffer, C.M. Andorf, T.E. Seigfried, D.A. Campbell, C.J. Lawrence, Choosing a genome browser for a Model Organism Database: surveying the Maize community, Database (Oxford) (2010) baq007.
[10] J.M. Nolan, V. Petrov, C. Bertrand, H.M. Krisch, J.D. Karam, Genetic diversity among five T4-like bacteriophages, Virol. J. 3 (2006) 30.
[11] T.G. Wolfsberg, Using the NCBI Map Viewer to browse genomic sequence data, Curr. Protoc. Hum. Genet. (2011) Chapter 18. Unit18 5.
[12] A. Grote, J. Klein, I. Retter, I. Haddad, S. Behling, B. Bunk, I. Biegler, S. Yarmolinetz, D. Jahn, R. Münch, PRODORIC (release 2009): a database and tool platform for the analysis of gene regulation in prokaryotes, Nucleic Acids Res. 37 (2009) D61–D65.
[13] S. Gama-Castro, H. Salgado, M. Peralta-Gil, A. Santos-Zavaleta, L. Muñiz-Rascado, H. Solano-Lira, V. Jimenez-Jacinto, V. Weiss, J.S. García-Sotelo, A. López-Fuentes, L. Porrón-Sotelo, S. Alquicira-Hernández, A. Medina-Rivera, I. Martínez-Flores, K. Alquicira-Hernández, R. Martínez-Adame, C. Bonavides-Martínez, J. Miranda-Ríos, A.M. Huerta, A. Mendoza-Vargas, L. Collado-Torres, B. Taboada, L. Vega-Alvarado, M. Olvera, L. Olvera, R. Grande, E. Morett, J. Collado-Vides, RegulonDB version 7.0: transcriptional regulation of *Escherichia coli* K-12 integrated within genetic sensory response units (Gensor Units), Nucleic Acids Res. 39 (2011) D98–D105.
[14] J. Baumbach, CoryneRegNet 4.0 — a reference database for Corynebacterial gene regulatory networks, BMC Bioinformatics 8 (2007) 429.
[15] W.J. Longabaugh, E.H. Davidson, H. Bolouri, Visualization, documentation, analysis, and communication of large-scale gene regulatory networks, Biochim. Biophys. Acta 1789 (2009) 363–374.